Referent Selection in Case-Crossover Analyses of Acute Health Effects of Air Pollution

Drew Levy,[1] Thomas Lumley,[2] Lianne Sheppard,[2,3] Joel Kaufman,[3,4] and Harvey Checkoway[1,3]

The case-crossover design was proposed for the study of a transient effect of an intermittent exposure on the subsequent occurrence of a rare acute-onset disease. This design can be an alternative to Poisson time series regression for studying the health effects of fine particulate matter air pollution. Characteristics of time-series of particulate matter, including long-term time trends, seasonal trends, and short-term autocorrelations, require that referent selection in the case-crossover design be considered carefully and adapted to minimize bias. We performed simulations to evaluate the bias associated with various referent selection strategies for a proposed case-crossover study of associations between particulate matter and primary cardiac arrest. Some *a priori* reasonable strategies were associated with a relative bias as large as 10%, but for most strategies the relative bias was less than 2% with confidence interval coverage within 1% of the nominal level. We show that referent selection for case–crossover designs raises the same issues as selection of smoothing method for time series analyses. In addition, conditional logistic regression analysis is not strictly valid for some case–crossover designs, introducing further bias. (Epidemiology 2001;12:186–192)

Keywords: Study design, data analysis, case-crossover design, bias, pollution.

The case-crossover design is suited to the study of a transient effect of an intermittent exposure on the subsequent risk of a rare acute-onset disease hypothesized to occur a short time after exposure. In the original development of the method,[1,2] effect estimates were based on within-subject comparisons of exposures associated with incident disease events with exposures at times before the occurrence of disease, using matched case-control methods or methods for stratified follow-up studies with sparse data within each stratum. The principle of the analysis is that the exposures of cases just before the event are compared with the distribution of exposure estimated from some separate time period. This distribution is assumed to be representative of the distribution of exposures for those individuals while they are at risk of developing the outcome of interest.

The health effects of fine particulate matter air pollution (PM) is a topical epidemiologic issue for which the case-crossover design may be especially useful. Fine particulate air pollution is an exposure that varies over time, and there is concern that PM may affect the incidence of acute cardiovascular and respiratory disease events.[3–5] Extensive time series of daily air pollution measures for metropolitan regions are often available for air pollution research. Most previous studies of the relation of air pollution and health events have been Poisson regression time-series analyses of health events.[6–8]

The use of alternative analytic approaches and statistical models may improve causal inferences about air pollution effects. In particular, when measurements of exposure or potential effect modifiers are available on an individual level, it is possible to incorporate this information into a case-crossover study unlike a time-series analysis. A disadvantage of the case-crossover design, however, is the potential for bias due to time trends in the exposure time-series.[9] Since case-crossover comparisons are made between different points in time, the case-crossover analysis implicitly depends on an assumption that the exposure distribution is stable over time (stationary). If the exposure time-series is non-stationary and case exposures are compared with referent exposures systematically selected from a different period in time, a bias may be introduced into estimates of the measure of association for the exposure and disease. These biases are particularly important when examining the small associations that appear to exist between PM and health outcomes.

[1]Department of Epidemiology, [2]Department of Biostatistics, [3]Department of Environmental Health, [4]Department of Medicine, University of Washington, Seattle, WA 98195–7232.

Address reprint requests to: Lianne Sheppard, Box 357232, Department of Biostatistics, University of Washington, Seattle, WA 98195–7232.

This research was funded in part by the United States Environmental Protection Agency through agreement CR825173–01–0 to the University of Washington and in part by the Health Effects Institute Research Agreement 97–2-2.

This research does not necessarily reflect the views of either funding agency and no official endorsement should be inferred.

Submitted April 25, 2000; final version accepted September 8, 2000.

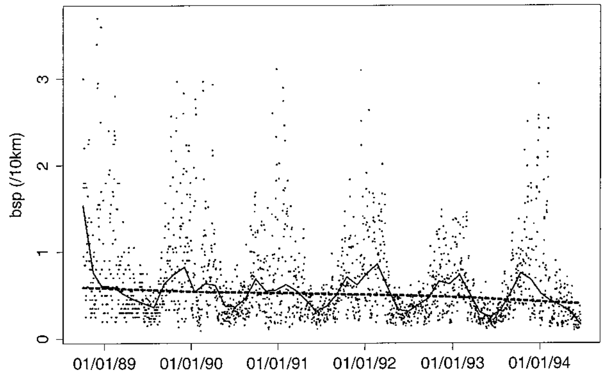

FIGURE 1. PM for study period showing long-term trend and seasonal pattern.

There are two ways in which the average PM level is not stationary over time, leading to possible bias. Long-term time trend occurs as pollution levels change gradually from year to year. When PM levels are increasing over time, *eg* from increased traffic, systematically selecting referents from an earlier period when pollution levels tend to be lower will give a positive bias; if PM levels are decreasing, *eg* owing to increased regulation, the bias will be negative. Figure 1 shows an example of a declining long-term trend in the Seattle PM data. In addition, there are distinct seasonal differences in PM levels. For example, Figure 1 shows that PM is highest in the winter, when rates of mortality and many forms of morbidity are higher for other reasons. Thus, long-term and seasonal trends may result in confounding.

Navidi[10] described an approach for addressing time-selection bias in case-crossover analyses — the ambidirectional case-crossover design — for exposures with time trends. When the occurrence of disease events does not affect subsequent exposure, as is the case with time-series of environmental exposures such as air pollution, Navidi proposed that all exposure times, before and after an index event, should be used as referents. By balancing referent exposures before the event with referent exposures that occur subsequent to the event, the time-selection bias due to linear time trend that occurs with unidirectional sampling is canceled out. Simulations by Bateson and Schwartz[11] show that the gross biases from seasonal variation can also be alleviated by choosing referents from a shorter period of time both before and after the case time.

Short-term (6 days or less) autocorrelation in PM time series is another concern. It is likely due to weather patterns that affect ambient PM concentration through source generation and accumulation in the atmosphere. Selecting referents from time adjacent to the case event times is conceptually similar to overmatching in conventional case-control studies.

This paper is based on simulations undertaken to explore the nature and degree of time-selection bias and to examine the ability of various strategies to counter biases anticipated in a case-crossover analysis of the association of fine particulate air pollution and out-of-hospital primary cardiac arrest. Simulation data are patterned on data from a population-based case-control study of 362 primary out-of-hospital sudden cardiac arrest that occurred in King County Washington from October 3, 1988 to June 25, 1994.[12] The strategies we considered are displayed graphically in Figure 2. The main strategy to be tested is that of using an ambidirectional referent sampling window restricted to 30 days before and after the occurrence of a case event (Figure 2C). Additionally, a 6-day window around the case event day excluding potential referent days is defined to address potential bias from short-term autocorrelation in the exposure time series (Figure 2D).

An alternative strategy was devised based on concern that short-term autocorrelation between the referents themselves may also be a source of bias in the estimates. The original strategy described above was elaborated further by requiring that there be a 6-day autocorrelation exclusion period between all observations used in the analysis. We believed that this requirement would allow for the necessary independence among all observations and would control for day-of-week effects on PM exposures. The alternative fixed interval strategies select referents only among lags (*ie* days before the index event) and leads (*ie* days after the index event) of 7, 14, 21, and 28 days (Figure 2E and 2F).

In addition to bias, statistical precision is important for analyses of health effects of air pollution that characteristically have small relative risks and a limited number of cases. Multiple referents may allow us to extract the maximum amount of information from the data. Since we can use pre-existing exposure time-series data, referents are cheap and optimal statistical efficiency dominates cost considerations. Our initial goal was to examine the variation in statistical precision that occurs as a function of the number of referents by exploring various analysis strategies. To our surprise, bias surfaced a the single dominant factor in the simulations. Thus, we focus on bias in this paper.

Methods

Data for particulate matter air pollution were obtained from the Puget Sound Clean Air Agency for October 3, 1988 through June 25, 1994. Daily averages of fine particulate matter air pollution as measured by nephelometer were used. The particle light scattering extinction coefficient (b_{sp}) measured by a nephelometer is an excellent proxy for daily variation in gravimetric measures of PM in the Seattle area, with a correlation of 0.94–0.95 between $PM_{2.5}$ and b_{sp} at three individual monitoring sites in the Seattle area.[13–15] We averaged observations from these three sites (Lake Forest Park, Duwamish, and Kent) to provide daily measures of exposure for the region. The study period had an average light scattering coefficient of 0.64×10^{-1} kilometers^{-1}. The range was .09 to 3.7 with values of 0.3, 0.47, and 0.81 at the first, second, and third quartiles. We permuted this data series to permit the evaluation of the effect of serial correlation on estimation by reordering the data by randomly sampling without replacement of the entire original data series.

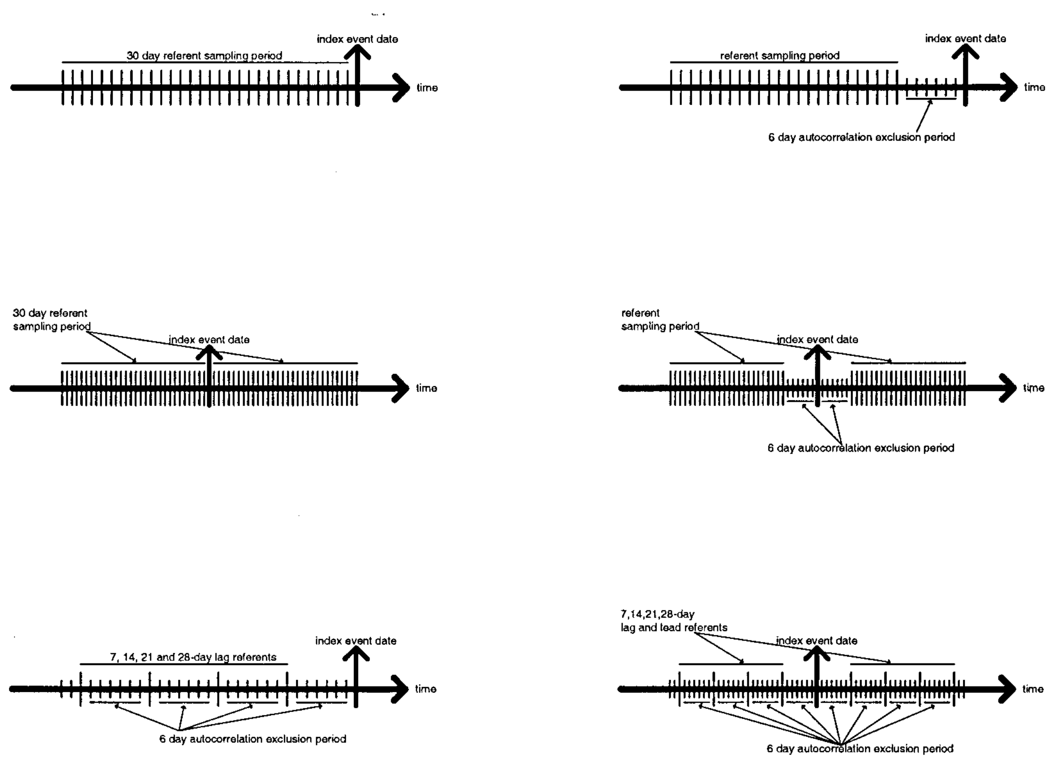

FIGURE 2. Referent selection strategies. (*A*) Series 2; retrospective referent selection within the 30-days prior to the case event. (*B*) Series 3; retrospective referent selection within the 30-days prior to the case event, with a 6-day autocorrelation exclusion period. (C) Series 4; ambi-directional referent selection. (*D*) Series 5; ambi-directional referent selection, with 6-day autocorrelation exclusion periods. (*E*) Series 8; retrospective fixed-interval referent selection. (*F*) Series 9; ambi-directional fixed-interval referent selection.

We simulated the occurrence of events. Reflecting the actual case series to be used in the analysis, 362 events were distributed over the 2,092-day time period as a function of exposure on day *j*. The probability that an event occurred at time t_j is given by the proportional hazards model,

$$\lambda = h(t, x) = h_0(t)e^{\beta x}, \qquad (1)$$

where the coefficient, β, was specified based on an incidence density ratio, $\exp(\beta)$, of 1.5 per interquartile range (IQR) change in b_{sp}:

$$\beta = \ln(1.5)/\text{IQR}.$$

This hazard ratio is larger than those previously observed in air pollution time series studies. It was anticipated[15] that for the specific, well-characterized outcome of primary cardiac arrest, the hazard ratio would be substantially larger than observed for more heterogeneous outcomes such as cardiovascular mortality. The interquartile range (IQR) for the CAB study period is $0.51 \times 10^{-1}\text{km}^{-1}\ b_{sp}$, so a hazard ratio of 1.5/IQR gives $\beta = 0.795$.

It is important to note that these simulations are not confounded by seasonal variation: outcome depends on time only through exposure. The simulations of Bateson and Schwartz[11] address seasonal confounding, and as we discuss below, it can also be addressed analytically.

Nine series of simulations were performed. The first series to assess the nature of time-selection bias used fixed retrospective single day lags: 365, 180, 90, 60, 30, 21, 14, 7, and 1 day before case events. Subsequent series were designed to assess the ability of various referent selection strategies to counter bias. All were restricted to selecting referents from some subset of the 30 days preceding and following the case event. The second series involved retrospective referent selection within the prior 30 days of the case event using 1, 2, 4, 10 randomly selected (with replacement) days, or all 30 days in the referent selection sampling frame (Figure 2A). The third series involved retrospective referent selection with the 6-day exclusion period using 1, 2, 4, 10 randomly selected days, or all 24 days in the referent selection sampling frame (Figure 2B). The fourth series involved ambidirectional referent selection without the 6-day exclusion period using 1, 2, 4, 10 randomly selected days, or all 60 days in the referent selection sampling frame (Figure 2C). The fifth series involved ambidirectional referent selection with the 6-day exclusion period using 1, 2, 4, 10 randomly selected days, or all 48 (61–13) days in the referent selection sampling frame (Figure 2D). The sixth and seventh series repeated the fourth and fifth series with permuted data. At the suggestion of a reviewer, a 10th simulation with data repermuted for each iteration was performed, using ambidirectional

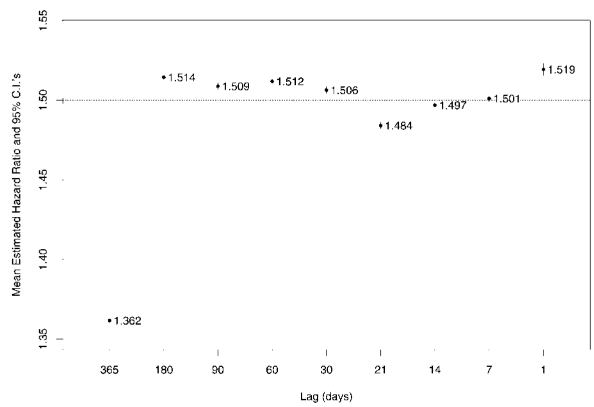

FIGURE 3. Series 1. Historical referent selection.

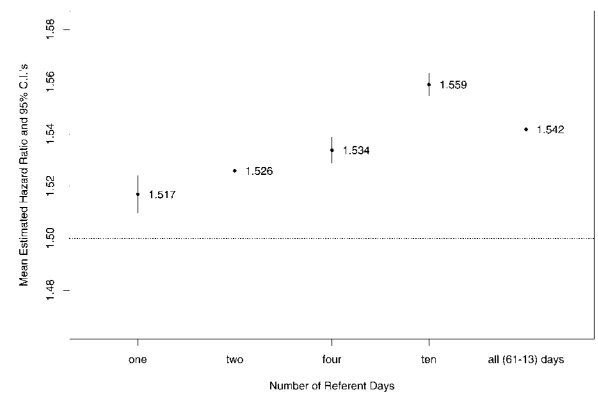

FIGURE 4. Series 5. Ambidirectional referent selection with 6-day exclusion.

referent selection without the 6-day exclusion period, using 10 randomly selected days. These permuted data series allow some examination of the role of trend and autocorrelation in the results.

The eighth and ninth series involved a referent selection strategy in which all cases and referents were required to be separated by 6 days within the ±30 day window. The eighth series retrospectively selected referents with 7; 7 and 14; 7, 14 and 21; and 7, 14, 21 and 28 day lags (Figure 2E). The ninth series ambidirectionally selected referents with 7; 7 and 14; 7, 14 and 21; and 7, 14, 21 and 28 day leads and lags (Figure 2F).

We calculated IQR relative risks and 95% confidence intervals for each iteration. We defined coverage as the percentage of simulations where the 95% confidence interval contained the true relative risk parameter. All analyses were performed in S-PLUS (MathSoft, Seattle). For the simulations of retrospective referent sampling at fixed lags, we did 10,000 iterations; all other simulations were iterated 1,000 times. Standard errors and confidence intervals for the mean of the individual estimated coefficients are based on the number of iterations.

Results

Here we concentrate on bias in case-crossover analyses. A more complete presentation of these simulations can be found in Levy.[14] Results are given for the coefficient estimate $\hat{\beta}$.

Figure 3 shows estimates of the association of PM and the incidence of events that occur when various single specific fixed lags are chosen to define the referent exposure. A lag of 1 year is associated with a negative 24% bias, with only 67% of the 95% confidence intervals for each of the estimates in the 10,000 iterations containing the value of the true coefficient. Lags of half a year through 1 month are biased in the range of 2.7 to 0.7%. A 21-day lag for referents is negatively biased by 2 to 3%, whereas lags 7 and 14 have biases of less than 1%. The 1-day lag is positively biased by 2.6–3.8%.

Choosing retrospective referent exposures from within a 30-day lag window shows a pattern of bias that is a function of the number of referents chosen (Table 1, series 2). A single referent is relatively unbiased. Choosing 10 referents randomly within the 30-day window improves the precision by one third (not shown), but is associated with a bias of 5 to 6%. Using all 30 days in the sample frame is associated with a bias of 3 to 4%. This pattern of increasing bias with the number of referents is exacerbated when the 6-day autocorrelation exclusion period is included in the definition of the sample frame (Table 1, series 3). With the 6-day exclusion, using 10 referents is associated with a positive bias of 12–14%.

Ambidirectional random sampling of referents within the ±30-day window results in a positive bias of 4% or less for all number of referents without the 6-day exclusion period (Table 1, series 4). For the ambidirectional series with the 6-day exclusion period (Table 1, series 5 and Figure 4), the bias begins in the range of 2% and increases with the number of referents to 9–10% with 10 referents. With the 6-day exclusion period, using all

TABLE 1. Multiple Referent Days: Percent Bias in $\hat{\beta}$ for Varying Numbers of Referents Sampled at Random from a Restricted Time Window

	Referent Sampling			Number of Referents				
Series	Timing	Excl.	Order	1	2	4	10	all
2	Retrospective	None	Orig.	−1.8–0.5	−0.5–1.5	2.1–3.8	4.7–6.1	2.7–4.0
3	Retrospective	6-day	Orig.	−1.7–0.6	1.9–3.8	6.6–8.3	12.3–13.8	11.6–13.0
4	Ambidirectional	None	Orig.	1.7–4.1	0.9–2.8	1.3–2.9	2.0–3.4	0.2–1.5
5	Ambidirectional	6-day	Orig.	1.6–3.9	3.3–5.2	4.7–6.3	8.8–10.2	6.1–7.4
6	Ambidirectional	None	Perm.	−0.1–1.7	0.3–1.7	0.7–1.8	1.7–2.5	−0.4–0.3
7	Ambidirectional	6-day	Perm.	0.1–1.9	0.2–1.6	0.8–1.9	2.7–3.6	0.6–1.3

Excl., exclusion; Orig., original; Perm., permutation.

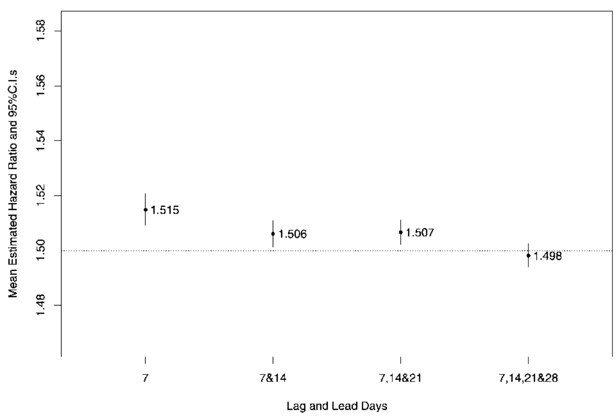

FIGURE 5. Series 9. Fixed-interval ambidirectional referent selection.

available days in the sampling frame as referents results in a bias of 6–7%, and yields coverage of only 88%.

Using permuted data to remove serial correlation from the time series results in biases in the range of −0.1–2% for 4 referents or less, regardless of whether the 6-day exclusion period is used (Table 1, series 6 & 7). For 10 referents, however, the bias is 1.7–2.5% without the ±6-day exclusion (series 6), and 2.7–3.6% with the ±6-day exclusion (series 7). Using all days in the window as referents in the permuted data yields an unbiased estimate without the ±6-day exclusion, and 0.6–1.3% bias with the ±6-day exclusion. Repermuting the data for each simulation yielded a bias of 2.0% for 10 controls, compared to 2.1% using a single fixed permutation.

Table 2 gives the bias estimates for the fixed-interval referent selection strategy. Both strategies show a monotonically changing pattern of bias. For the strategy limited to retrospective referents (series 8), the single 7-day lag is unbiased, but an increasing negative bias as large as −2.1 to −3.7% is evident as the 14, 21, and 28 day lags are included. For the strategy with ambidirectional referents (series 9 and Figure 5), the 7-day lead and lag is biased by 1.5–3.4%, and this bias progressively disappears as 14, 21, and 28 lead and lag days are added.

Discussion

These simulations focus on the nature of time selection bias and the effect of various schemes for referent selection in air pollution case-crossover analyses by contrasting three factors: 1) retrospective *vs* ambidirectional sampling, 2) the use of an exclusion to reduce short-term autocorrelation, and 3) the number of referents used. We also evaluated the influence of time-series patterns and serial correlation by permuting the data series. We observed complex patterns of bias in these simulations. We believe these complex patterns are the result of multiple competing sources of bias.

Overall, these simulations revealed distinct and sometimes substantial biases in most of the referent sampling strategies studied. While seeming sensible *a priori*, designing referent selection with restrictions specifically chosen to mitigate some forms of bias anticipated in analysis of PM time-series data would have been misguided (*eg* series 5). This approach could have led to bias in the range of 6–10% with 10 or more referents if employed in a naive analysis. Other plausible referent selection strategies produced entirely negligible bias. It is important to gain further insight into the sources of the biases observed, but it appears that they are likely to be relatively unimportant in practice.

Time-Selection Bias Patterns

The retrospective single fixed referent lag day series (series 1) reveals a non-monotonic pattern of bias. The extreme bias at the 365-day lag is qualitatively consistent with the expectation for the effect of a declining long-term time trend. If referents are systematically chosen from a period of time that tends to have higher exposure, then a bias toward the null is expected, as was observed. The positive bias of 2–3% seen for referent lags of 180, 90, and 60 days is qualitatively consistent with what might be expected for seasonal influences on referent exposure values, confounded to an unknown extent by the negative effect of long-term time trend. If cases tend to occur during the high air pollution seasons, choosing lags large enough to place referents in other seasons should make the referent exposures relatively lower. This selection bias would lead to the observed exaggeration of the estimated measure of association. The positive 1% bias seen at the 30-day referent lag suggests that there may be some small seasonal influence even at that proximity to the case event. The negative bias seen at the 21-day referent lag and, to a lesser extent, at the 14-day lag, indicates that for some unknown reason the referent exposures at those lags systematically tend to be greater than expected. This result may be related to cyclical weather patterns that influence local air pollution levels. The 7-day referent lag (seen also in series 8) seems to be unbiased while the 1-day referent lag shows a substantial positive bias. Overall, this complex pattern of biases indicates that there may be many patterns in the time series data that can influence effect estimation in various ways.

TABLE 2. Simulation Results: Percent Bias in $\hat{\beta}$ with Fixed Interval (Weekly) Referent Selection

Referent Sampling		Number of Lag or Lead/Lag Weeks			
Series	Timing	1	1 + 2	1−3	1−4
8	Retrospective	−1.3–1.1	−1.4–0.4	−2.0–−0.4	−3.7–−2.1
9	Ambidirectional	1.5–3.4	0.2–1.8	0.3–1.8	−1.0–0.4

Ambidirectional versus Retrospective Sampling

The original conception of the case-crossover design was retrospective since referents were chosen from times that preceded the event. This restriction is necessary when outcomes that may affect subsequent exposures are studied. In these situations, sampling referent times after event times could result in reverse-causation bias. For example, if exposure tended to decrease as a consequence of an event, using post-failure referent information could tend to bias risk estimates upward. The study of environmental exposure effects (as opposed to behavioral exposure effects) has the advantage that exposure levels subsequent to the event are unaffected by the event occurrence. Combined with a rare disease assumption this implies that one can use post-event exposures as referents.[10,11]

With the limitations of sampling within a narrow time interval, there is a tension between precision and bias. Restriction to a sampling window of 30 days serves the purpose of limiting the bias due to seasonality and coincidentally limits the influence of long-term time trend, which should be negligible in this relatively small window. A potential benefit of ambidirectional referent sampling in this context is to double the size of the sampling frame and permit a greater number of referents to be chosen. In the absence of bias, the greater number of referents provides a small but possibly valuable improvement in precision, as the variance of an estimator with m controls is proportional to $m/(m + 1)$. To maintain comparability between the designs, we did not take advantage of the opportunity to use more controls with ambidirectional sampling.

Conditional Likelihoods Used in Estimation Parameters for Matched Case-Control Studies

The fact that bias is still observed when the exposure series is permuted to remove any patterns shows that the bias cannot be entirely explained by seasonal variation or autocorrelation. A more careful mathematical analysis[16] shows that conditional logistic regression itself is inappropriate for these choices of referents. This phenomenon can be understood best in the context of a description of the conditional logistic regression estimation methods typically employed. The conditional likelihood[17] used in maximum likelihood estimation of regression parameters is based on comparing the case risk only to the risk of its referents. The conditional likelihood formula (L_C) reflects the probability of the observed data configuration relative to the probability of all possible permutations of the data configuration:

$$L_C = \frac{P_r(\text{observed configuration})}{P_r(\text{all possible configurations of the data})}. \qquad (2)$$

In an ideal matched case-control study, the controls are sampled independently and with equal probability from the stratum of the population containing the case. The observed configuration is that in which the case's exposure X_{i0} belongs to the case, and the other exposures X_{i1}, X_{i2}, ..., X_{im} belong to the m controls. This configuration is compared with all possible other configurations where, for example, X_{i2} is assigned to the case, and X_{i0}, X_{i1}, X_{i3}, ..., X_{im} to controls. With a single case in each matched set, the conditional likelihood in Eq 2 reduces to

$$\frac{\text{risk ratio for case}}{\text{sum of risk ratios for case and controls}}$$

where the risk ratio for an exposure X is $\exp(\beta X)$. When controls are not sampled at random from within the stratum of the case, this conditional likelihood may be invalid even in a matched case-control study. Austin *et al.*[18] show that choosing a friend or sibling as the matched control for each case can lead to bias. This bias is due to selection of controls from categories that are not mutually exclusive, leading to a situation in which the exposure of control subjects does not reflect that of members of the source population. The bias occurs because exposed individuals are either more or less likely than unexposed individuals to be included in multiple strata. Robins and Pike[19] consider this issue in more detail and give conditions for the conditional likelihood to be valid in a case-control study when controls are not sampled at random.

In some case-crossover designs, random sampling within mutually exclusive strata does not hold. For the design in which referents are taken at specific intervals before and after the case event with the case event in the center of the referent period, the observed configuration of the data is the only possible one. The same is true for the traditional case-crossover design in which the configuration with the referents before the case is the only one consistent with the study design. The true conditional likelihood in these designs would be identically equal to 1, so the likelihood used by conditional logistic regression is not the correct one. As our simulations show, the bias resulting from this incorrect conditional likelihood is small but not zero, and may in some circumstances be important. The bias does not exist for designs in which the case can occur at any point in the referent window, such as the original ambidirectional design proposed by Navidi,[10] but his design requires using the whole study time as the referent period, a requirement that in air pollution studies could produce confounding by season. A modified design that avoids seasonal confounding and gives the correct conditional likelihood can be obtained by dividing the time period *a priori* into fixed strata and using the remaining days in a stratum as referents for a case in that stratum. For example, calendar months could be used as strata; *eg* a case on Sunday, December 12, would be compared with all the other days in December. A finer stratification could be done on the day of the week as well, so that only the other Sundays in December would be used. In this situation, the positions of the cases are not determined by design, but instead vary randomly within the strata. We refer to these as *time-stratified* case-crossover designs. For further detail on this alternative, see Lumley and Levy.[16] It is also worth noting that this bias should be zero under the null hypothesis, at least in the absence of other covariates.[16]

Seasonal Confounding

Our simulations do not include any confounding by season, but still demonstrate sensitivity to the choice of referent periods. This sensitivity will be increased when confounding by season is present. The choice of referent period then presents a tradeoff between bias and variance in the estimated relative risks. In the presence of autocorrelation, there is less variation in exposure over a narrow referent window than over a wide one, and so for any fixed number of referents, the estimation of relative risks will be less precise for a narrower referent period. When a wider referent period is chosen, there is more information about the relative risk but this information is potentially contaminated by confounding.

The choice of referent period must be made based on substantive knowledge of the seasonal variability in pollution and mortality or morbidity in the area under study. This choice is the same as is made in a Poisson time series analysis when the amount of smoothing is chosen. In fact, in spite of claims that the two analyses are different,[20,21] a time-stratified case-crossover analysis without any individual-level time-varying covariates is mathematically equivalent to a Poisson time series analysis that uses dummy variables to estimate seasonal effects. Indicator variables are a crude but effective form of smoothing. Despite this equivalence the case-crossover design has the advantage of presenting the restriction in time in a way more familiar to epidemiologists.

Conclusions

The bias observed in this simulation study is predominantly positive. The nature of the bias may be different in other studies in which the autocorrelation in the time series, disease risk, correlation of events in time, and sampling designs may be different. In particular it should be noted that any variation in exposure between individuals will reduce the bias from improper conditioning, and that this bias would be negligible if different individuals had completely independent exposure histories (though other forms of bias might still be present).

For practical purposes this bias is relatively unimportant, and published ambidirectional case-crossover studies[20,21] that have chosen the referent days far enough apart to remove local autocorrelation should give reliable results (if our observations can be extrapolated) as demonstrated by the fixed interval strategy (series 9). It will often be indistinguishable from the finite sample bias inherent in estimating the regression parameter (compare Figures 3 and 4 in Lumley and Levy[16]). For future air pollution studies, modifying the design by partitioning the data *a priori* into mutually exclusive categories (true strata) rather than selecting potentially overlapping referent windows removes the bias completely,[16] providing a valid and elegant design.

Our simulation results show that three design features are useful for unbiased estimation. Unbiased estimation with conditional logistic regression requires dividing time into strata defined *a priori* and using the remainder of eligible days in each stratum as the referents for a case in that stratum, rather than selecting potentially overlapping referent windows centered at the time of each case event. If this approach cannot be taken, then seasonality and long-term time trend in the PM time series can be effectively removed by restriction of the sample frame for referents to a period short enough to be free of significant seasonal transitions. Furthermore, restricting the referent sampling window to require a 6-day interval between all exposures (the third design feature) ensures independence among observations and, more important in practice, controls for day of the week effects.

The latter two design features will also help control confounding by season and by day of the week. This confounding was not present in our simulations but will be important in real case-crossover analyses of air pollution data, and could cause much larger biases than those we have shown.

References

1. Maclure M. The case–crossover design: A method for studying transient effects on the risk of acute events. Am J Epidemiol 1991;133:144–153.
2. Mittelman MA, Maclure M, Robins JM. Control sampling strategies for case-crossover studies: an assessment of relative efficiency. Am J Epidemiol 1995;142:91–98.
3. Schwartz J. What are people dying of on high air pollution days? Environ Res 1994;64:26–35.
4. Dockery DW, Pope CA. Acute respiratory effects of particulate air pollution. Annu Rev Public Health 1994;15:107–32.
5. Schwartz J. Air pollution and daily mortality: a review and meta-analysis. Environ Res 1994;15:107–132.
6. Vedal S. Ambient particulates and health: lines that divide. J Air Waste Management Assoc 1997;47:551–581.
7. Schwartz J, Spix C, Touloumi G, Bacharova L, Barumandzadeh T, Tertre AL, Piekarksi T, Ponce de Leon A, Ponka A, Rossi G, Saez M, Schouten JP. Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions. J Epidemiol Community Health 1996;50(suppl 1): S3–S11.
8. Samet JM, Zeger SL, Berhane K. The association of mortality and particulate air pollution. In: Particulate Air Pollution and Daily Mortality: Replication and Validation of Selected Studies. Cambridge, MA: Health Effects Institute, 1995;1–104.
9. Greenland S. Confounding and exposure trends in case-crossover and case-time-control designs. Epidemiology 1996;7:231–239.
10. Navidi W. Bidirectional case-crossover designs for exposures with time trends. Biometrics 1998;54:596–605.
11. Bateson TF, Schwartz J. Control for seasonal variation and time trend in case-crossover studies of acute effects of environmental exposures. Epidemiology 1999;10:539–544.
12. Siscovick DS, Raghunathan TE, King I, Weinmann S, Wicklund KG, Albright J, Bovbjerg V, Arbogast P, Smith H, Kushi LH, Cobb LA, Copas MK, Praty BM, Retzlaff BM, Childs M, Knopp RH. Dietary intake and cell membrane levels of long-chain N-3 polyunsaturated fatty acids and the risk of primary cardiac arrest. JAMA 1995;274:1363–1367.
13. Sheppard L, Levy D, Norris G, Larson TV, Koenig JQ. Effects of ambient air pollution on nonelderly asthma hospital admissions in Seattle, Washington, 1987–1994. Epidemiology 1999;10:23–30.
14. Levy D. A Case-Crossover Study of Particulate Matter Air Pollution and Out-of-Hospital Primary Cardiac Arrest in King County, Washington. PhD thesis, University of Washington, Seattle, WA, 1999.
15. Levy D, Sheppard L, Checkoway H, Kaufman J, Lumley T, Koenig J, Siscovick D. A case-crossover analysis of particulate matter air pollution and out-of-hospital primary cardiac arrest. Epidemiology 2001;12:193–199.
16. Lumley T, Levy D. Bias in the case-crossover design: implications for studies of air pollution. Environmetrics 2000;11:689–704.
17. Breslow NE, Day NE. Statistical Methods in Cancer Research. vol. 1. The Analysis of Case-Control Studies. IARC Scientific Pub. No. 32. Lyon: International Agency for Research on Cancer, 1980;192-246..
18. Austin H, Flanders WD, Rothman KJ. Bias arising in case-control studies from selection of controls from overlapping groups. Int J Epidemiol 1989; 18:713–716.
19. Robins J, Pike M. The validity of case-controls studies with nonrandom selection of controls. Epidemiology 1990;1:273–284.
20. Neas LM, Schwartz J, Dockery D. A case-crossover analysis of air pollution and mortality in Philadelphia. Environ Health Perspect 1999;107:629–631.
21. Lee JT, Schwartz J. Reanalysis of the effects of air pollution on daily mortality in Seoul, Korea: A case-crossover design. Environ Health Perspect 1999;107:633–636.